SAFE AI AT CARIAD

Peter SCHLICHT

Head of "Safe AI and AI Innovations »,
CARIAD (Volskwagen)

Safe Artificial Intelligence for Automotive Software

Peter Schlicht | Head of Safe Artificial Intelligence @CARIAD

CARIAD

01

Introduction CARIAD

02

Approaching Safe AI within ADAS and AD

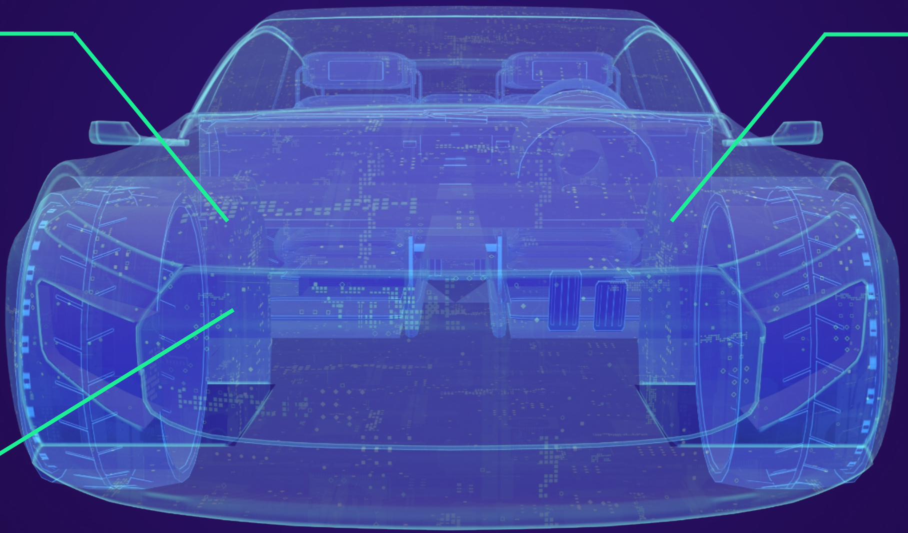
03

Challenges for Safe AI

04

Opportunities

The automotive industry is in the midst of a software revolution - a beginning of a new era



New tech capabilities

- // Artificial intelligence
- // Virtual & augmented reality
- // Quantum computing

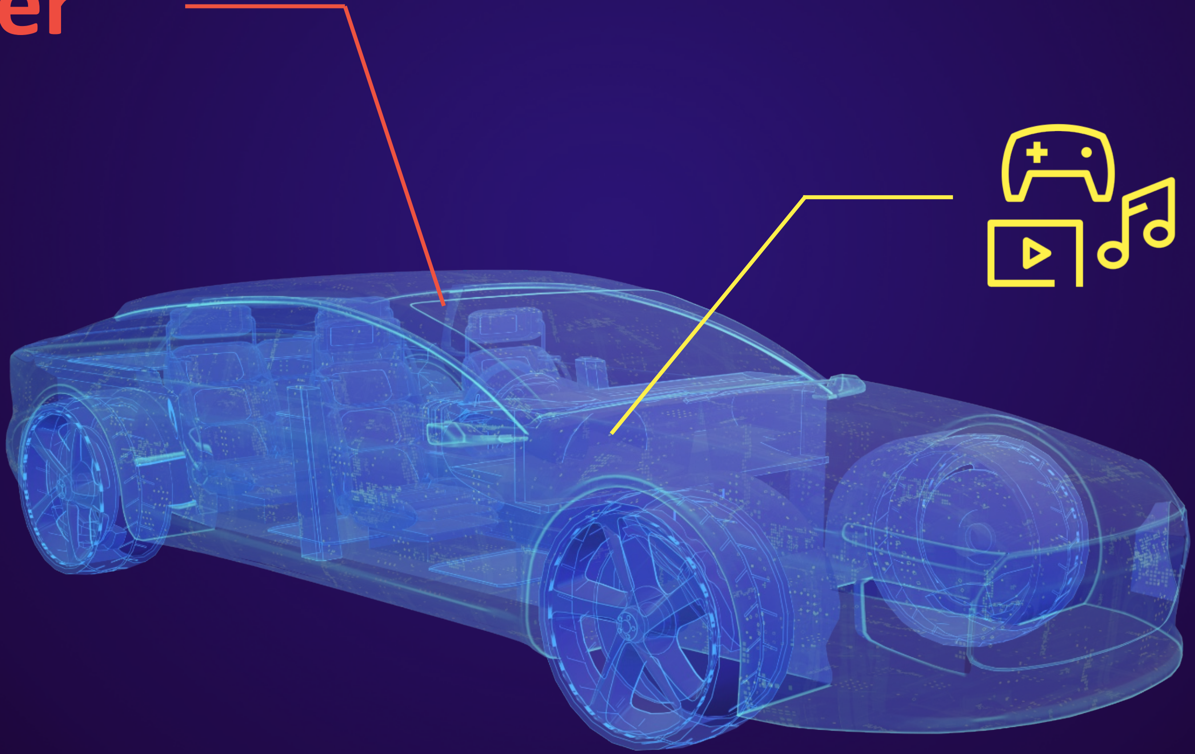
New revenue streams

- // Function on demand
- // Updates & upgrades
- // Data & e-commerce

New customer expectations

- // Assisted & automated driving
- // Always up-to-date over lifetime
- // Immersive user experience

CARIAD makes automotive mobility safer, more comfortable and more sustainable…



The 'Driver’

RELAX & BE SAFE

The Digital Experience

ENJOY THE RIDE & STAY CONNECTED

… and enables the future of the Volkswagen Group brands

Updatability

Constant updates for the best, always fresh customer experiences.

Speed

Seamless software platform and intelligent data analysis speed up development and time to market.

Scalability

One platform from entry-level to top-end getting better each day thanks to large amounts of data collected by VW Group fleet.

Simplicity

One unified platform reduces complexity.

Customer orientation

Data-oriented development helps to learn from and react to customers' needs.

New revenue streams

Enabling new digital business models: From after-sales to monetizing third-party apps.

01

Introduction
CARIAD

02

Approaching Safe
AI within ADAS
and AD

03

Challenges for
Safe AI

04

Opportunities

Safety for ADAS means combining multiple streams

Evidence Provision for safety argument

Safety arguments will rely on a **safety concept, mitigation** techniques and **test evidence**.

Evidence generation / validation

Validation of Components

Within development: need for validation reuse (e.g. open loop component tests, safety performance indicators)

Reusable closed-loop testing

Handling the building blocks of driving – mostly on object level

Can be improved by simulation of perception failures

Validation in fleets

Testing the full stack in Shadow most and campaign-based testing both in real-world traffic and on proving grounds

Argument of residual risk

Statistically sound performance estimates

Development according to SOTA

Mitigation techniques for known failures

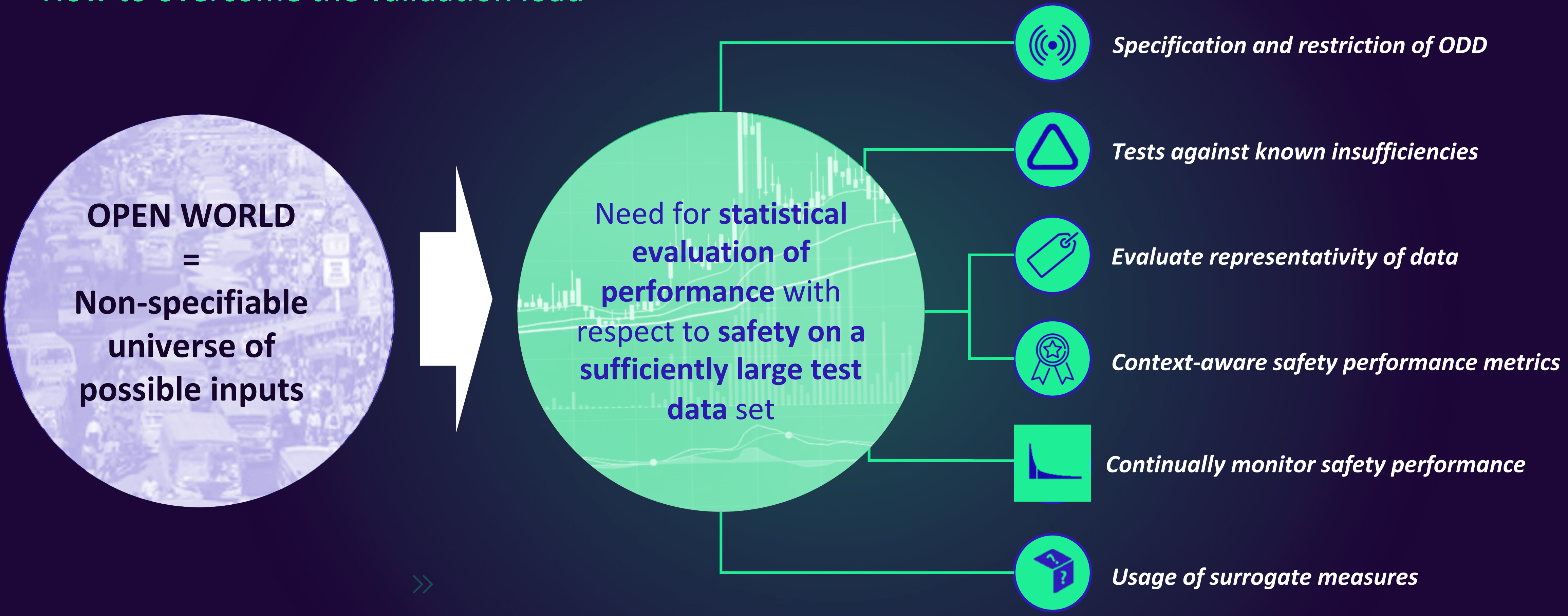
System Design allowing for resilience

Using SW design paradigms to raise Safety by design.

Focusing design decisions on resulting safety performance.

Mechanisms to reduce validation needs

How to overcome the validation load



OPEN WORLD = Non-specifiable universe of possible inputs

Need for **statistical evaluation of performance** with respect to **safety on a sufficiently large test data** set

- *Specification and restriction of ODD*
- *Tests against known insufficiencies*
- *Evaluate representativity of data*
- *Context-aware safety performance metrics*
- *Continually monitor safety performance*
- *Usage of surrogate measures*

Systematic understanding and data-driven approach to validation will lead to manageable efforts

Embracing uncertainties

How to deal with the big unkown

Many sources of uncertainty for Safe AI and Safe Systems

- Changing regulations (EU AI Act, ISO PAS 8800)
- Continually changing ODD
- Rare corner cases
- Rapid technological progress in AI
- Incremental development

Safety and Safe AI are developed incrementally as well

- Perfomance estimation and definition of safety-mechanisms in a data-driven fashion
- Derivation of situational performance requirements based on overall performance

Systematic understanding and data-driven approach to validation will lead to manageable efforts

01

Introduction CARIAD

02

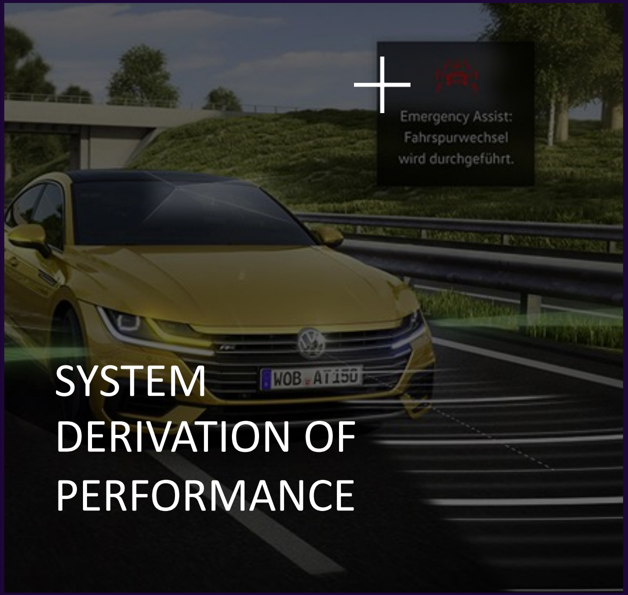
Approaching Safe AI within ADAS and AD

03

Challenges for Safe AI

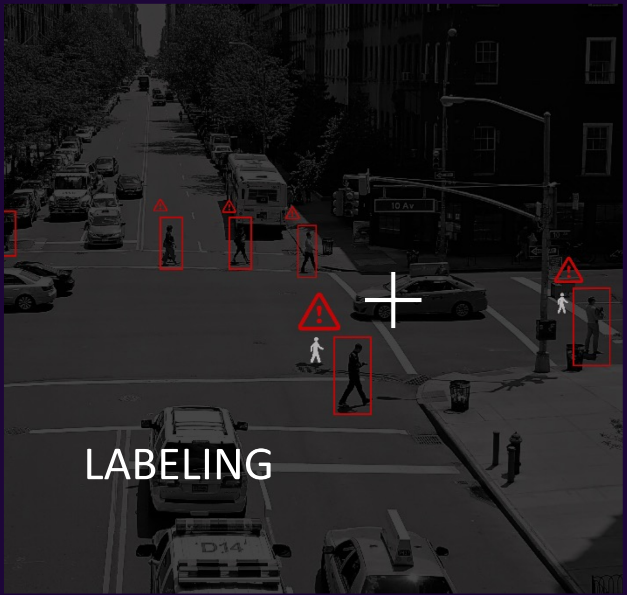
04

Opportunities



SYSTEM DERIVATION OF PERFORMANCE

ROBUSTNESS



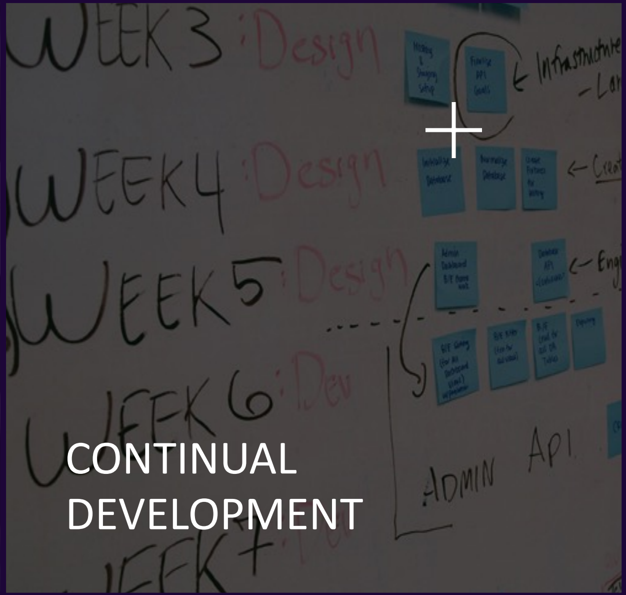
LABELING

HANDLING COMPLEXITY

ORGANISATIONAL EXCELLENCE

RAPID INCREMENTS

QUALIFICATION AND TRACEABILITY



CONTINUAL DEVELOPMENT

Testing Data for DNNs in Automated Driving

The relevance of test data samples changes over time

Shifting Label Class Limits

Contextually Different Assessments

New Classes Appearing

Distributional Shifts, Domain Drifts

Classes change there appearance

Behavioral changes of dynamic objects

Testing Data for DNNs in Automated Driving

The relevance of test data samples changes over time

- **Shifting Label Class Limits**
- Contextually Different Assessments
- New Classes Appearing
- Distributional Shifts, Domain Drifts
- Classes change there appearance
- Behavioral changes of dynamic objects

Testing Data for DNNs in Automated Driving

The relevance of test data samples changes over time

- **Shifting Label Class Limits**
- Contextually Different Assessments
- New Classes Appearing
- Distributional Shifts, Domain Drifts
- Classes change there appearance
- Behavioral changes of dynamic objects

01

Introduction
CARIAD

02

Approaching Safe
AI within ADAS
and AD

03

Challenges for
Safe AI

04

Opportunities

Generative AI

Foundational Models and Intelligent analysis for rapid data driven safety

- Foundational models contain strong semantic world understanding and scale over multiple tasks
- Generative AI can be used to automatically create both development artifacts (requirements, test cases etc) as well as test data
- Foundational models can be used for automatic tagging, ingestion of safety knowledge and analysis of large amounts of evidence

Recent AI forthcomings will revolutionize Safety Engineering

Cooperation across industries, academia, society

Reaching a reflected concensus

- An approach to safety will ultimately be dependent on choices and assumptions that need careful weighing and argumentation
- Cross-society cooperation will lead to accepted standards
- Regulation and standardization will need to go hand in hand with academic process
- A rapid transition of innovation to industrial application will serve performance and safety.

Through cooperation, progress is accelerated considerably.

CARIAD